YIN LIN

irenelin@umich.edu
https://niceirene.github.io/

EDUCATION

University of Michigan, Ann Arbor   Ann Arbor, MI
Ph.D. student in Computer Science and Engineering   Sept. 2019 – Dec. 2024 (Expected)
Advisor: Prof. H. V. Jagadish

Shanghai Jiao Tong University   Shanghai, China
B.S. in Computer Science   Sept. 2015 – June 2019

RESEARCH INTEREST

Responsible data management; efficient data wrangling and lineage; pipeline optimization; AI4DB

PUBLICATIONS

Conferences

- *Mitigating Subgroup Unfairness in Machine Learning Classifiers: A Data-Driven Approach*
  Yin Lin, Samika Gupta, H. V. Jagadish
  ICDE - 40th IEEE International Conference on Data Engineering, 2024.

- *SMARTFEAT: Efficient Feature Construction through Feature-Level Foundation Model Interactions*
  Yin Lin, Bolin Ding, H. V. Jagadish, Jingren Zhou
  CIDR - The Conference on Innovative Data Systems Research, 2024. *[link]*

- *Predicate Pushdown for Data Science Pipelines*
  Cong Yan, Yin Lin, Yeye He
  SIGMOD - Proceedings of the ACM on Management of Data 1 (2), 1-28, 2023. *[link]*
  Awarded SIGMOD 2023 Best Paper Award

- *On Detecting Cherry-picked Generalizations*
  Yin Lin, Brit Youngman, Yuval Moskovitch, H. V. Jagadish, Tova Milo
  PVLDB - Proceedings of the VLDB Endowment, 15(1): 59-71, 2022. *[link]*

- *Identifying Insufficient Data Coverage in Databases with Multiple Relations*
  Yin Lin, Yifan Guan, Abolfazl Asudeh, H. V. Jagadish
  PVLDB - Proceedings of the VLDB Endowment, 13(11): 2229-2242, 2020. *[link]*

- *R^2-Tree: An Efficient Indexing Scheme for Data Center Networks*
  Yin Lin, Xinyi Chen, Xiaofeng Gao, Bin Yao, Guihai Chen
  DEXA - International Conference on Database and Expert Systems Applications, 2018. *[link]*

Workshops

- *OREO: Detection of Cherry-picked Generalizations*
  Yin Lin, Brit Youngman, Yuval Moskovitch, H. V. Jagadish, Tova Milo
  PVLDB - Proceedings of the VLDB Endowment, 15(12): 3570 - 3573, 2022. *[link]*

- *On Structural vs. Proximity-based Temporal Node Embeddings*
  Puja Trivedi*, Alican Büyükçakır*, Yin Lin, Yinlong Qian, Di Jin and Danai Koutra
  KDD MLG - International Workshop on Mining and Learning with Graphs, 2020. *[link]*

- *MithraDetective: A System for Cherry-picked Trendlines Detection*
  Yoko Nagafuchi, Yin Lin, Kaushal Mamgain, Abolfazl Asudeh, H. V. Jagadish, You (Will) Wu, Cong Yu
  CoRR, 2020, arXiv/2010.08807. *[link]*

Journal Articles

- *Representation Bias in Data: A Survey on Identification and Resolution Techniques*
  Nima Shahbazi, Yin Lin, Abolfazl Asudeh, H. V. Jagadish
  ACM Computing Surveys. *[link]*

- *A Survey on Techniques for Identifying and Resolving Representation Bias in Data*
  Nima Shahbazi, <u>Yin Lin</u>, Abolfazl Asudeh, H. V. Jagadish
  CoRR, 2022, arXiv/2203.11852. *[link]*

PROFESSIONAL EXPERIENCE

Alibaba Group, Bellevue,
Research Intern, in Data Analytics and Intelligence Lab (DAIL) (Mentor: Bolin Ding)   May 2023 – Aug. 2023

- Leveraged pre-trained large language models (LLMs) for feature enhancement through operator-guided search and code generation, enabling efficient application of LLMs to data tasks.

- Added safeguards to prevent the generation results that create unhelpful features, which could degrade prediction performance.

Microsoft Research, Redmond,
Research Intern in Data Management, Exploration and Mining (DMX) (Mentor: Cong Yan)   May 2022 – Aug. 2022

- Developed a search-verification approach for predicate pushdown to optimize the execution efficiency of data science pipelines with non-relational operators and user-defined functions (UDFs).

- Applied predicate pushdown for tracing row-level data lineage, achieving an easily adaptable approach and efficient lineage querying with extensive operator coverage.

University of Michigan, Ann Arbor,
Graduate Student Research Assistant (Advisor: Prof. H. V. Jagadish)   Since Sept. 2019

- Thesis: Exploring Intersectionality in Responsible Data Management.

University of Waterloo, Waterloo
Research Intern in Software Architecture Group (Advisor: Prof. Meiyappan Nagappan)   July 2018 – Oct. 2018

TECHNICAL SKILLS

Responsible AI, Database Systems, Machine Learning Fairness, LLM for Data Manipulation, Data Provenance
Programming: Python, C++, SQL, Java, JavaScript

HONORS & AWARDS

NSF Travel Award for ICDE 2024	2024
SIGMOD best paper award	2023
Rackham Dean's and Named PhD fellowship *Full first-year PhD fellowship from the University of Michigan*	2019
Outstanding Undergraduate in Shanghai Jiao Tong University	2019
Academic Scholarship *Awarded to top 10% undergraduates for academic performance at SJTU*	2016 – 2018
SCSK Scholarship *Awarded to 7 computer science undergraduates at SJTU for academic performance*	2018
Yitu Scholarship *Awarded to 3 computer science undergraduates at SJTU for their research*	2017
Huawei Scholarship *Awarded to 7 computer science undergraduates at SJTU for academic performance*	2017
Chun Tsung Scholar *Awarded to 50 undergraduates at SJTU, funded by Nobel Prize owner Tsung-Dao Lee*	2016

PROGRAM COMMITTEES

NeurIPS: Conference on Neural Information Processing Systems, 2024
CIKM: ACM International Conference on Information and Knowledge Management, 2023, 2024
AIBSD: AAAI Workshop on Artificial Intelligence with Biased or Scarce Data, 2024
ReLM: AAAI Workshop on Responsible Language Models, 2024
ICDE: IEEE International Conference on Data Engineering, 2024 (external)